NBER WORKING PAPER SERIES

EXPECTATIONS FORMATION WITH FAT-TAILED PROCESSES: EVIDENCE AND THEORY

Tim de Silva
Eugene Larsen-Hallock
Adam Rej
David Thesmar

Working Paper 33808
http://www.nber.org/papers/w33808

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2025

We are grateful to Xavier Gabaix for very helpful feedback, and thank Mauro Cazzaniga for his outstanding research assistance. We also thank Nick Barberis, Eben Lazarus, Stefan Nagel (discussant), Kelly Shue, and seminar participants at CFM, Columbia, MIT, Yale, NYU, University of Toronto, Stanford, Wharton, and the NBER Behavioral Finance Meeting for their thoughtful comments. Thesmar is a consultant for Capital Fund Management. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Expectations Formation with Fat-Tailed Processes: Evidence and Theory
Tim de Silva, Eugene Larsen-Hallock, Adam Rej, and David Thesmar
NBER Working Paper No. 33808
May 2025
JEL No. D84, D91, G41

ABSTRACT

This paper studies expectations formation when the underlying process has fat tails. Using a large sample of firm sales growth expectations, we document three facts: (i) the relationship between forecast revisions and future forecast errors is strongly non-linear, (ii) the distribution of sales growth has fat tails, and (iii) extreme values of sales growth tend to mean-revert. We formally show that these three facts are consistent with a model in which the underlying process is non-Gaussian, but forecasters fail to recognize this fully. We estimate this model and show it quantitatively explains our three facts. Finally, we show the model is consistent with evidence from an online forecasting experiment where the underlying process is non-Gaussian and the non-linearity in the momentum of stock returns.

Tim de Silva
Stanford University
Graduate School of Business
Department of Finance,
and Stanford Institute for Economic Policy Research (SIEPR)
tdesilva@stanford.edu

Eugene Larsen-Hallock
Capital Fund Management
eugene.larsen-hallock@cfm.com

Adam Rej
Capital Fund Management
Adam.REJ@cfm.com

David Thesmar
Massachusetts Institute of Technology
MIT Sloan School of Management
and NBER
thesmar@mit.edu

Expectations formation is a core question in economics. In recent years, a strand of literature in macroeconomics and finance has been collecting empirical regularities using survey data on subjective forecasts. A common finding in this literature is that forecast errors are predictable using information within forecasters' information sets, which is inconsistent with rational expectations. Most theories proposed to explain this predictability assume that the data-generating process (DGP) being forecasted is simple, such as an AR1 process with normal shocks, and forecasters know this DGP. However, despite knowing the DGP, forecasters make predictable errors because of behavioral biases or cognitive limitations, such as sluggish updating (Bouchaud et al., 2019), representativeness (Bordalo et al., 2019, 2020a), and imperfect memory (Afrouzi et al., 2023).

In this paper, we take a different approach and argue that recognizing the presence of fat tails in the underlying DGP is crucial for understanding the properties of subjective forecasts. We begin by documenting three facts using data on sales growth forecasts by equity analysts: (i) the relationship between forecast revisions and future forecast errors—the variables used in Coibion and Gorodnichenko (2015) regressions—is strongly non-linear; (ii) the distribution of the underlying process has fat tails; and (iii) the conditional expectation of future sales growth is non-linear in current growth, with mean reversion in the tails. Next, we build a forecasting model that connects these facts. The key ingredients in our model are that the underlying process is non-Gaussian, but forecasters fail to recognize this. After showing formally that our model can explain the three facts we documented in the data, we estimate it and show that it does so quantitatively. Finally, we show that our framework is consistent with evidence from an online forecasting experiment where the underlying process is non-Gaussian and that it provides an explanation for non-linearity in the momentum of stock returns.

Our empirical analysis uses data on analyst forecasts of sales growth from IBES. The advantage of focusing on sales growth rather than earnings-per-share (as is typically done) is that sales growth is stationary and provides a larger sample. Using these data, the first and most important fact that we document is that the relationship between forecast revisions and future forecast errors is strongly non-linear. In some papers, revisions linearly and *positively* predict forecast errors, a feature commonly interpreted as evidence of underreaction (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019). In others, revisions linearly and *negatively* predict forecast errors, which is evidence of overreaction (Bordalo et al., 2019, 2020a). In our panel of forecasts and realizations of sales growth, we show that both co-exist. For intermediate values of revisions, forecasters underreact to news (a positive relationship between revisions and errors). For large values of revisions, forecasters overreact (a negative relationship between revisions and errors). Additionally, we show that this non-linear relationship between errors and revisions is driven by cross-sectional rather than aggregate variation, and does not vary with analysts' forecasting experience. This

suggests that our error-revision relationship is unlikely to be driven by a slow convergence of learning that can occur with one short time-series (Bianchi et al., 2022; Farmer et al., 2024).

Our second empirical fact is that the distribution of sales growth has fat tails, as in Stanley et al. (1996). To detect the presence of fat tails, we examine the log-density plots, similar to the literature on income dynamics (Guvenen et al., 2021). In the bulk of the sales growth distribution, we find that a Gaussian density, which is quadratic in logs, provides a close approximation. However, in the top and bottom 10% of the distribution, sales growth has much thicker and longer tails than those of a normal distribution. Instead, we find that these tails are well-approximated by a power law with a tail parameter between 2 and 3, thinner than Zipf's law (Gabaix, 2009). We show that these fat tails do not arise from heterogeneity in volatility across firms (Wyart and Bouchaud, 2003), consistent with Moran et al. (2024), and do not arise from time-varying aggregate volatility.

Our third and final motivating fact is that the conditional expectation of future sales growth on current growth is non-linear. In particular, we find that the relationship between current and one-year-ahead sales growth is increasing and linear in the bulk of the distribution. However, in the tails of the distribution, we find that this relationship flips sign, indicating that extreme values of sales growth tend to mean-revert. As described below, this non-linearity helps inform the exact way the DGP in our model deviates from that of standard Gaussian models.

In the next part of the paper, we develop a forecasting model that is designed to connect the above three facts. In the model, the DGP for the forecasting variable contains a persistent and transitory component, which is common in models of dividend growth (Bansal and Yaron, 2004; Lettau and Wachter, 2007) and income dynamics (Guvenen et al., 2014). Importantly, we assume the transitory component is non-Gaussian and follows a power law distribution, a tractable way of characterizing processes with fat tails (Gabaix, 2009). In contrast, we assume that the persistent component follows an AR1 process with normal innovations. This assumption is less crucial, but it is important that the transitory component has thicker tails than the persistent one. Having fat tails only in the persistent component would not match our third fact: the strong mean reversion in sales growth.

We show that this model of the DGP is consistent with our second and third facts. The model generates fat tails by assumption, consistent with our second fact. To show that it generates the non-linear conditional expectation of future growth conditional on current growth—our third fact—we leverage a result from empirical Bayes theory known as Tweedie's formula (Efron, 2012). Although this expectation cannot be expressed in closed-form due to the non-normality, this result allows us to characterize it as a function of the (observable) density function of growth. We show that this

result implies that the expectation of future growth conditional on current growth is locally linear when the density is locally Gaussian, as in the bulk of the distribution. However, when the density is locally non-Gaussian, as in the tails, the expectation of future growth conditional on current growth is no longer linear and is asymptotically decreasing in current growth. Intuitively, very large realizations of growth are likely due to the transitory component and, hence, will not likely persist.

Given this model of the DGP that is consistent with our second and third facts, we show it can also explain the non-linear relationship between forecast errors and revisions—our first fact—with a single assumption: agents construct their forecasts ignoring fat tails. Formally, we assume that agents form forecasts according to the Kalman filter, which would be the rational expectation in our model if the transitory component followed a normal distribution rather than a power law. We view the assumption that agents use a simple mis-specified model as a natural form of bounded rationality (as in Fuster et al. 2010 and Gabaix 2019), but do not microfound it. We show that this assumption is enough to generate overreaction in the tails and underreaction in the bulk. This is because large revisions are driven by large shocks to current growth, which asymptotically come from the transitory component of the DGP. While a rational forecaster would recognize that these extreme shocks are unlikely to persist, our agents that ignore fat tails do not and, therefore, overreact. However, because our agents are unbiased unconditionally, overreaction in the tails has to be compensated by underreaction in the bulk, consistent with our first fact.

Next, we assess our model's ability to replicate our three facts quantitatively. We estimate the parameters governing the DGP using simulated minimum distance and show that they provide a close fit to our second and third facts that are specific to the DGP. When we turn to our first fact, we find that the estimated model generates too much underreaction in the bulk and overreaction in the tails. This is not surprising, given our model has no additional free parameters that govern belief formation. Therefore, we enrich our model by assuming that expectations are a weighted average of the Kalman filter and the rational expectation (as in Fuster et al. 2010 and Gabaix 2019). We estimate the degree of shrinkage towards the rational expectation, which we numerically compute using a particle filter (Fernandez-Villaverde and Rubio-Ramirez, 2007), and find that the model can match our first fact with forecasters placing a 71% weight on the rational expectation and a 29% weight on the Kalman filter. Having computed the rational expectation, we compare its accuracy with that of the Kalman filter and find that it is relatively small. This supports the idea that it would be a "default" in a model of bounded rationality (Fuster et al., 2010; Gabaix, 2019).

We conclude with two additional tests of our theory that forecasters do not fully recognize the presence of fat tails. First, we run an online forecasting experiment similar to Afrouzi et al. (2023), but where the underlying process has fat tails. The key benefit of this experiment is that it

allows us to test our theory of expectation formation directly by experimentally varying the features of the data-generating process. When we run the experiment using our estimated DGP, we find that the relationship between errors and revisions is non-linear, like in our first fact. In contrast, when forecasters forecast a similar process with no fat tails, we find no evidence of a non-linear relationship between errors and revisions. These findings provide direct evidence that the non-linear relationship between errors and revisions in the data is driven by the fat tails of the DGP, which is the key result of our theory.

Second, we show that our model makes predictions for return momentum (Jegadeesh and Titman, 2011) that are supported by the data. To translate sales growth expectations into returns, we apply the Campbell (1991) return decomposition with a constant subjective discount rate (as in Bouchaud et al. 2019 and Nagel and Xu 2019). Our model predicts that the relationship between past and future returns should be positive in the bulk of the distribution, where underreaction to news is dominant. In contrast, it predicts mean-reversion of returns in the tails, where agents fail to recognize that the extreme shocks are not persistent. We find support for this prediction in the universe of smaller stocks: for these stocks, momentum tends to mean-revert for extreme losers and winners.

Related literature. This paper contributes to the recent and growing literature on expectations formation in two ways. First, we contribute to the empirical literature that has documented under and overreaction across many forecasting variables and horizons. Broadly speaking, this literature tends to find evidence of underreaction when looking at shorter-term or consensus forecasts (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019), and overreaction when looking at longer-term or individual forecasts (Bordalo et al., 2019, 2020a; Wang, 2021). Relative to this empirical literature, our contributions are to provide field evidence of both under and overreaction within the *same* forecasting variable and horizon, and lab evidence that the degree of overreaction can vary within a sample depending on the Pareto tail in the DGP.

Second, we contribute to the literature that proposes models of belief formation that can generate under and overreaction, including constant-gain learning (Nagel and Xu, 2019), selective recall (Bordalo et al., 2020b), and biased perceptions of autocorrelation (Wang, 2021). Most of this literature works with models in which data-generating processes are Gaussian, implying conditional expectations are linear. In contrast, our empirical analyses highlight how this relationship can be quite non-linear, and we provide a theory that links this non-linearity to the presence of under and overreaction within a given process. Our focus on a non-Gaussian dynamics is similar to Kozlowski et al. (2020), but we focus on forecasters ignoring these dynamics rather than learning. With forecasts of a single time series, rational learning can converge relative slowly (Farmer et al., 2024), generating an in-sample relationship between errors and revisions even under full rationality

(Singleton, 2021; Bianchi et al., 2022). In contrast, one benefit of having the large cross-section in our data is that, under fairly mild assumptions, it averages out the portion of this predictability that is driven by a short time series. Our assumption that forecasters (partially) ignore non-Gaussian dynamics is inspired by the literature on bounded rationality, which argues that economic agents use simplified models to minimize computation costs (Fuster et al., 2010; Gabaix, 2019). However, it is possible that this assumption could be microfounded via Bayesian learning about the tail parameter of the process.

Three closely related papers are Kwon and Tang (2025), Augenblick et al. (2024), and Graeber et al. (2024). Kwon and Tang (2025) also provide a model of belief formation with non-Gaussian dynamics. In their model, news events belong to categories with different power law distributions, and forecasters have diagnostic expectations, causing them to overreact to news from categories with fatter tails and underreact to news from categories with thinner tails. This prediction is similar to our model. An important difference is that this provides a theory of why over- and underreaction would vary depending on the category from which a realization is drawn, while our model provides a theory of why over- and underreaction would vary even *within* a category. Augenblick et al. (2024) propose a model in which forecasters incorrectly perceive signal quality and shrink it to a default, which leads to overreaction to weak signals and underreaction to strong ones. As we discuss in the paper, our model could be interpreted in this way. Given our model of the DGP, past growth is a much stronger predictor of future growth in the bulk of the distribution rather than in the tails. Our forecasters that ignore non-Gaussian dynamics do not realize this, causing them to underreact to strong signals and overreact to weak signals. Finally, Graeber et al. (2024) analyze the S-shaped relationship between returns and earnings surprises, which loosely map into revisions and forecast errors. In their theory, the strong sensitivity between returns and surprises around zero comes from overreaction that occurs at a category boundary, while the lower sensitivity away from zeros comes from dampening within a category due to noisy perceptions. Our first empirical fact—underreaction in the bulk and overreaction in the tails—points to a different mechanism in our setting.

Through our model of the DGP with fat tails, we connect the expectations formation literature with the literatures on power laws. The omnipresence of power laws (Gabaix, 2009) suggests that the misperception of fat tails that we document is likely important for understanding subjective forecasts in other settings. The facts that we document about the data generating process of sales growth are consistent with the literature on firm dynamics. Our second fact that sales growth (rather than its level) has fat tails was first documented by Stanley et al. (1996) and recently emphasized by Boar et al. (2025).[1] Our third fact is consistent with Jaimovich et al. (2025), who find that revenue

[1] Our finding that the heavy tails of the growth distribution cannot be explained by a mixture of Gaussian distribution with heterogeneous variances is also consistent with Moran et al. (2024).

is more persistent in the bulk of the distribution than in the tails and exhibits fat tails. Also related is the literature on income dynamics, which emphasizes deviations from the canonical income processes with Gaussian shocks (Guvenen et al., 2014, 2021). However, this literature on income dynamics emphasizes the importance of non-normal persistent shocks, while the key ingredient in our model is non-normal *transitory* shocks.[2]

Outline. Section 1 describes our data source, variable construction, and sample selection. Section 2 documents our three main facts. Section 3 lays out the simple framework we build to connect and explain these facts. Section 4 estimates our model and shows that it quantitatively explains these facts. Section 5 provides additional tests of our model from an online return experiment and data on stock returns.

1 Data, Variable Construction, and Sample Selection

1.1 Data Source

Our analysis primarily relies on a large annual panel of analyst forecasts for yearly revenues at one and two-year horizons. We obtain these forecasts from IBES Adjusted Summary Statistics files, which provide data for both U.S. and international firms. The summary statistics files contain "current" estimates as of the third Wednesday of each month. We extract mean forecasts reported in the third month of each fiscal year $t+1$, after fiscal year t earnings have been announced. The forecasts we use correspond to two horizons: fiscal years $t+1$ and $t+2$. The resulting panel covers the period 2000-2023, and includes both U.S. and foreign firms.

We focus on revenue (i.e., sales) forecasts rather than earnings-per-share (EPS) forecasts for several reasons. First, revenues are consistently positive, making past realized revenues a natural normalization base (some normalization is needed to ensure stationarity). Revenue growth forecasts exhibit a well-behaved distribution with minimal outliers—a crucial attribute given our emphasis on distribution tails. Second, revenues are not reported on a per-share basis, eliminating the confounding effects of unexpected stock splits that can generate substantial jumps in EPS or forecast errors unrelated to the focus of this paper. Finally, the distribution of EPS-to-price ratios—a common normalization for most studies—is notably non-normal with a characteristic bulge above zero, whereas the log sales growth distribution demonstrates smooth, well-behaved properties.

[2]Another difference is that our focus is primarily on generating the kurtosis in the data rather than skewness. The latter is a pervasive feature of income data due to extreme negative events like job loss.

In our robustness checks, we also examine individual analyst forecasts and EPS forecasts (normalized by price). Similar to our revenue data, we extract these at annual horizons in the third month of the fiscal year.

1.2 Definition of Forecasting Variable: Sales Growth

We start with the definition of sales growth, which is our forecasting variable of interest. We denote sales of firm i at date t as R_{it}. These data are from the IBES actual files, which ensures comparability with analyst forecasts (described below). (Raw) log sales growth is given by $G_{it} = \log R_{it} - \log R_{it-1}$.

Most of our analysis actually focuses on *adjusted* sales growth. We do this for two reasons. First, this adjustment makes different firms comparable with one another, which makes it easier to fit a single data-generating process on the whole cross-section of firms (we do this in Section 4). Second, as discussed in Wyart and Bouchaud (2003), the thick tail of the growth distribution may mechanically emerge from the combination of normal processes interacted with heterogeneous variances. Our adjustment takes care of these two issues, but we will later explore robustness.

For each firm i, denote T_i as the number of years for which we have a growth observation and $\mu_i = \frac{1}{T_i}\sum_t G_{it}$ as the empirical average of growth observations for firm i. Additionally, denote $\sigma_i = \frac{1}{T_i}\sum_t |G_{it} - \mu_i|$ as the mean absolute deviation, an estimate of the standard deviation of growth at the firm level. The advantage of this measure is that it is less sensitive to outliers than variance, because it has no squared term. We then define *adjusted* growth as:

$$g_{it} = \frac{G_{it} - \mu_i}{\sigma_i}.$$

1.3 Definitions of Forecast Errors and Revisions

For each firm i and each year t, we denote $F_t R_{it+1}$ the forecast made in year t for the future realization of sales R_{it+1}, where $F_t R_{it+1}$ is obtained from IBES summary files as the mean consensus forecast extracted in the third month after the end of fiscal year t. Similarly, the two-year ahead forecast $F_{t-1}R_{it+1}$ is measured three months after the end of fiscal year $t-1$.

Our key variable of interest in this paper is the forecast of log sales growth. We construct the

h-year ahead forecast of *raw* sales growth as

$$F_t G_{it+h} = \log F_t R_{it+h} - \log F_t R_{it+h-1},$$

where $F_t R_{it} = R_{it}$. The h-year forecast of *adjusted* sales growth is then

$$F_t g_{it+h} = \frac{1}{\sigma_i} \left(\log F_t G_{it+h} - \mu_i \right).$$

We focus on $h = 1$ and $h = 2$ for one- and two-year ahead forecasts. Note that the way we translate from forecasts of raw to adjusted sales growth implicitly ignores a Jensen's inequality term because $\log F_t G_{it+h} \neq F_t \log G_{it+h}$. We ignore this adjustment for simplicity, but it does not materially affect our analysis. First, on a theoretical level, a constant Jensen's term would simply shift the unconditional level of forecasts, while our analysis focuses on the conditional properties of forecasts errors. Second, we show that our empirical results are robust to using percent (instead of log) growth.

Using these definitions, we then construct forecasts errors and revisions in line with the literature on expectations formation (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019). In particular, raw and adjusted forecast errors are defined as

$$ERR_t G_{it+1} = G_{it+1} - F_t G_{it+1}, \quad ERR_t g_{it+1} = g_{it+1} - F_t g_{it+1},$$

while raw and normalized forecast revisions are defined as

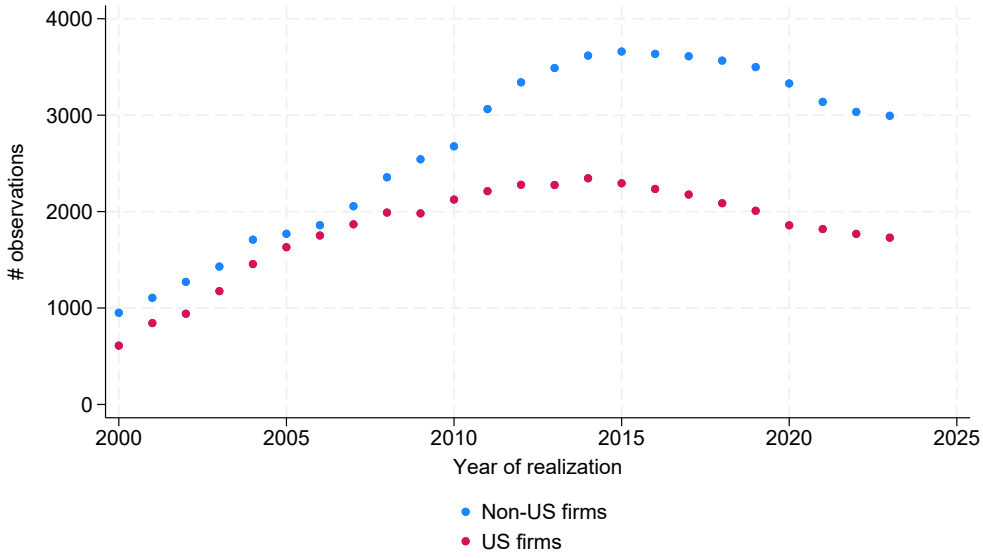
$$R_t G_{it+1} = F_t G_{it+1} - F_{t-1} G_{it+1}, \quad R_t g_{it+1} = F_t g_{it+1} - F_{t-1} g_{it+1}.$$

1.4 Sample Selection and Summary Statistics

The sample in our analysis consists of the entire international IBES summary file, subject to several sample restrictions. First, we focus on firms for which at least 10 realizations of sales growth are observed (i.e., $T_i \geq 10$). Second, we focus on years for which sales forecasts are sufficiently well-populated, which is from 2000 to 2023. We end up with 122,395 observations. We then trim the top and bottom 1% of all of our variables. This trimming procedure does not materially affect our results, but discards extreme outliers which would compress our plot axes.

Figure 1 shows the total number of observations in our data for which we have a non-missing growth and one-year forecast. As early as 2000, there are already 1,700 such observations (1,000

Figure 1: Number of observations in our sample



Notes: This figure shows the total number of observations in our data for which we have a non-missing (adjusted) growth and one-year forecast.

international and 700 in the U.S.). Then, the number of firms grows to about 3,700 internationally and 2,200 in the U.S. Overall, sales growth is well covered by IBES after 2000.

Table 1 shows summary statistics for our main variables of interest. The standard deviation of raw log sales growth is 17 ppt, and the mean is 6%. Normalized growth has a mean and median closer to zero, and a standard deviation of 1.20 – it is not exactly equal to 1 because we normalize by the mean absolute deviation, not the firm-level standard deviation. Forecast errors have less variance than growth itself, consistent with the idea that forecasts contain some information about future growth. Figure A.1 provides suggestive evidence that the distribution of forecast errors has thicker tails than a Gaussian fit.

2 Motivating Facts

In this section, we document the three key facts that motivate the model of expectations formation that we develop in Section 3. In the main body of the paper, we present these facts using normalized log sales growth, g_{it}, but discuss a series of robustness checks and report their results in Appendix.

Table 1: Summary Statistics

Variable	Mean	P25	Median	P75	SD	# of Obs.
Panel A: Full Sample						
Log sales growth (raw)	0.06	-0.02	0.06	0.14	0.17	108,706
Growth forecast error (raw)	-0.01	-0.05	-0.00	0.05	0.13	108,694
Growth forecast revision (raw)	-0.01	-0.04	-0.00	0.03	0.11	105,609
Log sales growth (normalized)	0.00	-0.70	0.01	0.72	1.20	108,706
Growth forecast error (normalized)	-0.06	-0.52	-0.02	0.45	0.96	108,694
Growth forecast revision (normalized)	-0.04	-0.39	-0.03	0.32	0.80	105,609
Panel B: US Firms						
Log sales growth (raw)	0.06	-0.02	0.05	0.13	0.17	64,881
Growth forecast error (raw)	-0.01	-0.06	-0.00	0.05	0.13	64,686
Growth forecast revision (raw)	-0.00	-0.04	-0.00	0.03	0.10	63,892
Log sales growth (normalized)	0.01	-0.70	0.02	0.73	1.20	64,727
Growth forecast error (normalized)	-0.07	-0.59	-0.04	0.47	1.01	64,569
Growth forecast revision (normalized)	-0.02	-0.36	-0.01	0.34	0.79	63,716
Panel C: Non-US Firms						
Log sales growth (raw)	0.07	-0.01	0.06	0.15	0.18	43,825
Growth forecast error (raw)	-0.01	-0.05	-0.00	0.05	0.12	44,008
Growth forecast revision (raw)	-0.01	-0.05	-0.01	0.03	0.12	41,717
Log sales growth (normalized)	-0.00	-0.70	-0.01	0.72	1.20	43,979
Growth forecast error (normalized)	-0.03	-0.43	-0.01	0.41	0.88	44,125
Growth forecast revision (normalized)	-0.07	-0.43	-0.05	0.30	0.81	41,893

Notes Source: IBES summary files. Raw growth corresponds to unadjusted log growth (G_{it} in the main text). Adjusted growth subtracts the firm-level mean and divides by the mean absolute distance (g_{it} in the main text). All variables are trimmed at the bottom and top 1%.

2.1 Fact #1: Non-Linear Relationship Between Forecast Errors and Revisions

Our first and central empirical fact is based on a regression of forecast errors on forecast revisions. This regression was introduced by Coibion and Gorodnichenko (2015) (CG) and takes the following form:

$$ERR_t Y_{it+1} = \alpha + \beta R_t Y_{it+1} + e_{it+1} \tag{1}$$

for any forecasting variable Y_{it}. This regression is useful because its slope coefficient can be used to distinguish between different models of expectations formation, requiring only panel data on expectations. Full-information rational expectations predicts $\beta = 0$ for consensus forecasts, while limited-information rational expectations predicts $\beta = 0$ for individual forecasts. When this regression is run using consensus forecasts, $\beta > 0$ is typically interpreted as evidence of information frictions, as in models of sticky or noisy information (Coibion and Gorodnichenko, 2015). However, with individual forecasts, $\beta > 0$ is interpreted as non-Bayesian underreaction (Bouchaud et al., 2019), while $\beta < 0$ is interpreted as overreaction (Bordalo et al., 2020a). A key feature of prior

literature is that it restricts analysis to linear functional forms, as in equation (1). While this is a natural starting point, especially in settings with small sample sizes, in this section we use our large sample of sales expectations to provide evidence that this relationship is non-linear.

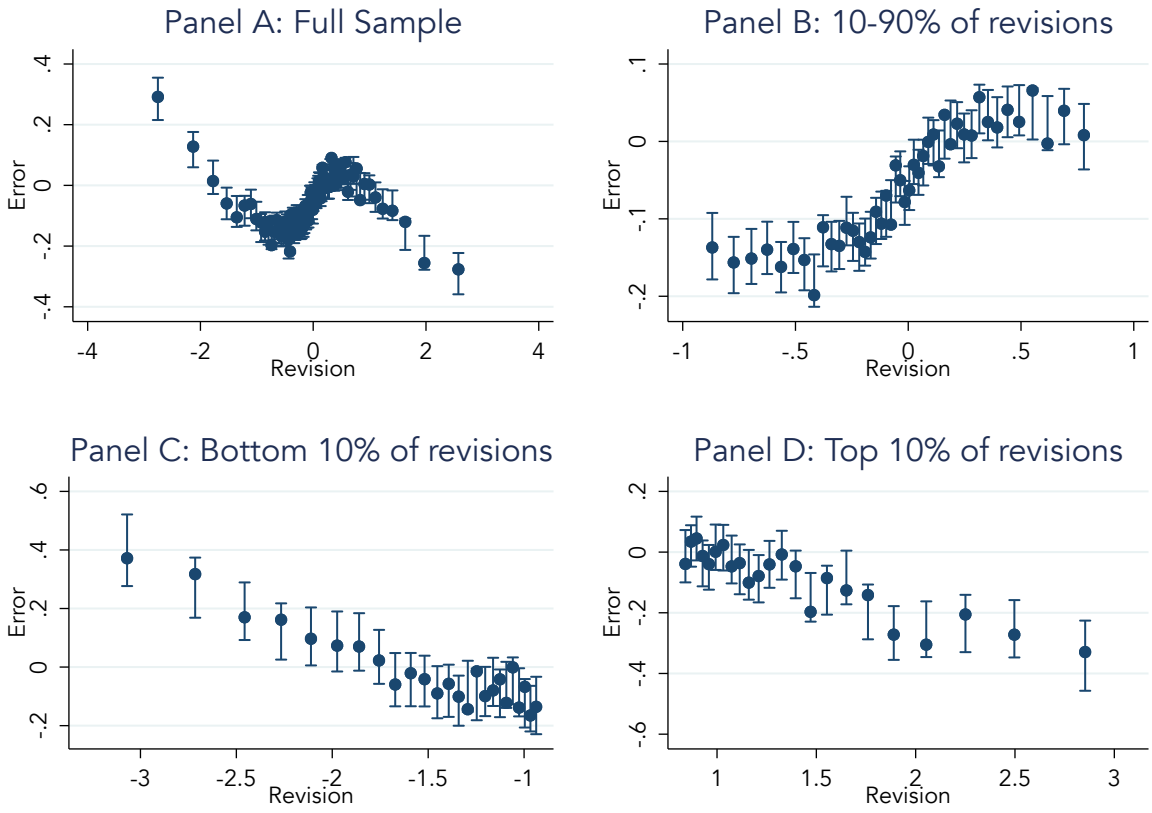
Figure 2 provides preliminary evidence that, for $Y = g_{it}$, the relationship between forecast errors and revisions is non-linear. Panel A shows a binned scatterplot of forecast errors, $ERR_t g_{it+1}$, as a function of revisions $R_t g_{it+1}$, using 100 bins. This plot provides evidence of significant non-linearity. Panels B, C, and D show the same plot separately for observations in the 10-90th, 0-10th, and 90-100th percentiles of revisions. For revisions in the bulk of the distribution, Panel B shows that errors are increasing in revisions, consistent with forecasters underreacting to news that causes moderately-sized revisions in forecasts. This finding is not novel to our paper: it is consistent with the underreaction in analysts' EPS forecasts for US firms (Bouchaud et al., 2019), as well as managers' revenue forecasts in the US and Italy (Ma et al., 2020).

Panels C and D of Figure 2 shows that for large (positive or negative) revisions, the positive relationship between forecast errors and revisions reverses and becomes negative: large positive (negative) forecast revisions are predictive of negative (positive) future forecast errors. Unlike the relationship in the bulk of the distribution, which is consistent with underreaction, this finding is consistent with *over*-reaction. In other words, forecasters appear to overreact in response to news that generate large revisions, while underreacting to more moderate news.

Figure 3 provides sharper statistical evidence of the non-linear relationship between forecast errors and revisions shown in Figure 2. In particular, we report the slope coefficient from estimating the CG error-revision regression on eight different subsamples based on different percentiles of revisions shown in the horizontal axis. In each one of these regressions, we double cluster error terms at the year and firm levels. The results in Figure 3 confirm the findings in Figure 2: between the 20th and 80th percentiles of revisions, errors are a significantly *increasing* function of revisions. In contrast, outside of these bounds in the tails of the distribution of revisions, errors are *negatively* correlated with revisions.

Robustness. The non-linear relationship between errors and revisions that we document is robust to a battery of robustness checks. First, while our adjustments from raw to normalized (log) growth are done to adjust for heterogeneous variances, it could be that this adjustment mechanically generates mean-reversion. We show in Figure A.2 that this is not the case: raw (log) growth displays the same non-linear pattern. Second, working with the natural logarithm of growth has the property of somewhat compressing the tails of the distribution. However, Figure A.3 shows that percent growth exhibits the same non-linear relationship between errors and revisions. Third, we attempt to

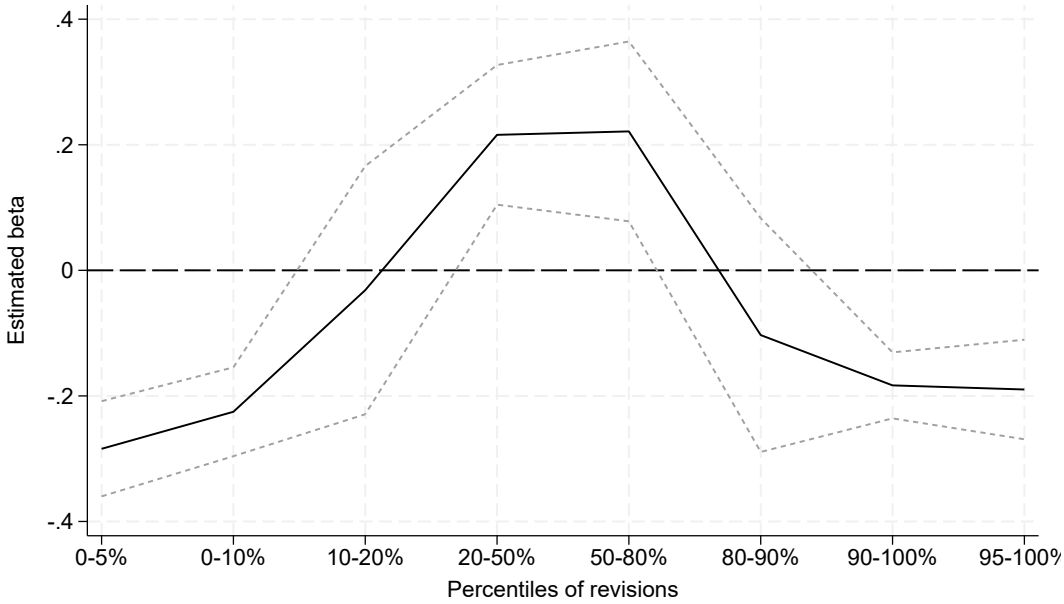
Figure 2: Non-Linear Relationship Between Forecast Errors and Revisions



Notes: This figure shows binned scatterplots of forecast errors on forecast revisions for normalized log sales growth. See definitions of raw and normalized growth in Section 1. Panel A shows the results for the entire sample; Panel B restricts the sample the 10-90th percentiles of revisions; Panel C restricts the sample to below the 10th percentile of revisions; Panel D restricts the sample to above the 90th percentiles of revisions. Vertical bars represent 95% confidence intervals, assuming the relationship is piecewise linear and continuous (option "ci(1 1)" in stata command "binsreg").

control for aggregate shocks by making a different adjustment to growth. Each year, we compute the cross-sectional mean absolute deviation of raw log growth (from that year's mean log growth) as a measure of aggregate dispersion. We then divide raw error and raw revision by this measure of dispersion. Figure A.4 shows that, after such adjustment for time-varying aggregate volatility, the non-linear relationship between errors and revisions is still very strong. Fourth, because we seek to maximize the size of our sample, our data include non-US firms and therefore somewhat differs from most existing research. We show in Figure A.5 that the non-linear pattern is clearly present both in US and international firms separately. Fifth, forecasting noise has been suggested as a potential source of downward bias in the CG coefficient (de Silva and Thesmar, 2024). To assess the importance of noise, Figure A.6 reimplements the exercise on individual analyst forecasts, which likely contain more noise than the consensus. Consistent with the presence of noise, we find

Figure 3: Error-Revision Regression Coefficient by Percentiles of Revisions



Notes: In this Figure, we report the estimates of β in the following regression:

$$ERR_t g_{it+1} = \alpha + \beta R_t g_{it+1} + \epsilon_{it+1}$$

where g_{it} is the normalized log sales growth rate defined in Section 1. This regression is run on eight different subsamples, whose ranges are described in the x-axis of this chart. These subsamples corresponds to the tails and the bulk of the distribution of revisions. The point estimate of β is the solid black line, while the dashed lines corresponds to the 95% confidence interval based on standard errors that are double-clustered by firm and year.

the intermediate positive slope, is smaller and the tail negative slopes are more negative. However, the non-linear relationship between errors and revisions is still very strong, suggesting that noise is unlikely to explain our main fact.

Part of the literature on expectations formations has emphasized that rational learning can occur slowly (Farmer et al., 2024), which can generate a correlation between errors and revisions in short time series that may not be present out of sample (Bianchi et al., 2022). Our use of cross-sectional variation alleviates this concern to the extent that our results are driven by firm-specific variation, but not if they are driven by aggregate variation. Figure A.7 shows that are results are indeed driven by the former: our results are unchanged after absorbing aggregate shocks with time fixed effects. Additionally, we directly test whether learning drives the non-linear error revision relationship by splitting our sample into four quartiles of analyst experience, measured by the number of firms followed by each analyst up to the current year. Figure A.8 shows that the non-linear pattern is quantitatively similar in all quartiles. This finding suggests that learning is not a first-order driver of

our results, which is likely driven by our use of extensive cross-sectional variation.

Finally, while the present paper focuses on sales growth because of its empirically convenient properties, most literature on analyst forecasts analyzes EPS forecasts. In Figure A.9 we show that raw EPS forecast exhibit the same type of non-linearity, albeit slightly less pronounced (it is also present in Bouchaud et al. 2019).[3]

Overall, the evidence on forecast errors and revisions points towards a different treatment of large versus smaller shocks. Such evidence is hard to square with established models of expectations formations, which feature linear DGPs (typically, AR1 models) and linear expectations models. We will deviate from the existing literature in allowing for fat tails in the growth process. To guide our theory, we first document two additional facts on the fat tails of firm dynamics.

2.2 Fact #2: Fat Tails in Distribution of Sales Growth

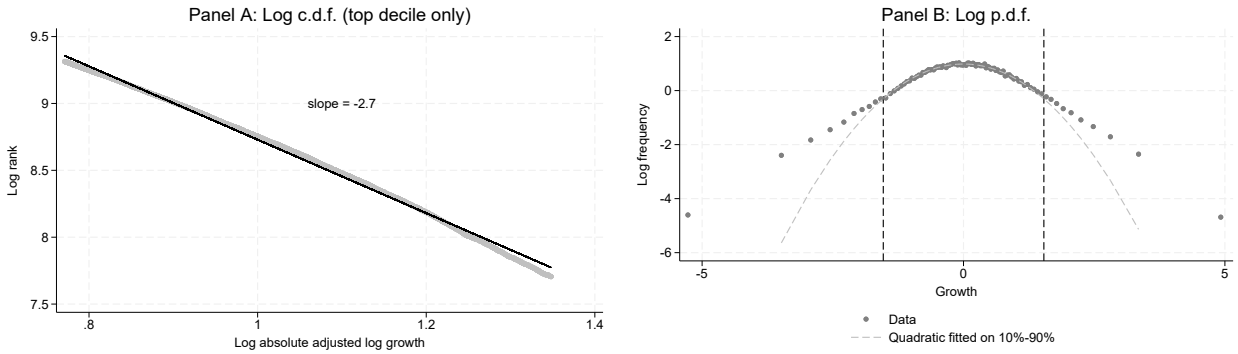
Our second fact is that the distribution of sales growth has fat tails, which informs the specification of the data-generating process in our model. It is well-known that many financial and economic variables have tails that are fatter than a normal distribution Gabaix (2009). For example, the distribution of firm sizes follows a Pareto distribution with a tail coefficient of one (Axtell, 2001), which is known as Zipf's law, and the distribution of growth rates in COMPUSTAT follows a Laplace distribution, which has fatter tails than a normal distribution (Stanley et al., 1996; Bottazzi and Secchi, 2006).

Figure 4 examines the tails of log normalized growth, g_{it}, in our sample. Panel A shows a classic plot in the literature on power laws: a plot of log rank against $\log |g_{it}|$ in the tail of the distribution, where each points corresponds to an observation (Gabaix, 2009). In this plot, we focus on the top 10% of observations of absolute growth $|g_{it}|$, and exclude the top 1%. The plot also shows the OLS regression line, which has a slope of -2.7. The fact that the relationship between these two variables is approximately linear shows that in the top decile of the distribution, the density function of g_{it} is well-approximated by a power law of tail coefficient 2.7 (it has a variance but no kurtosis).

Another way to illustrate the fat tails in the distribution of sales growth is shown in Panel B of Figure 4. This plot shows the log of the probability density function of g_{it} for the entire distribution. We compute this density by first grouping observations into centiles of sales growth. For each

[3]To normalize EPS, we use the standard practice of dividing by a common lagged value of stock price, as in de Silva and Thesmar (2024). Hence, the forecast error is defined as $\frac{EPS_{it+1}-F_{jt}EPS_{it+1}}{P_{it-2}}$ for analyst j, firm i at date t. Similarly, revisions are defined as $\frac{F_{jt}EPS_{it+1}-F_{jt-1}EPS_{it+1}}{P_{it-2}}$.

Figure 4: CDF and PDF of Sales Growth Distribution



Notes: Panel A of this figure shows a scatterplot of log rank of $|g_{it}|$ against $\log|g_{it}|$. We restrict ourselves to the top decile of absolute growth and remove the top one percent. The panel reports the slope of the regression of log rank on log growth estimated by OLS. Panel B shows the log density of g_{it} computed as follows. For each centile, we estimate density as the log of the number of observations in the centile divided by its range. The dashed line is a quadratic fit on the centiles between the 10th and 90th centiles. The two dashed vertical lines correspond to the cutoff values of the top and bottom decile of the distribution.

centile, we then compute the average growth, which is shown on the x-axis. On the y-axis, we calculate the density as the difference between the log frequency in the centile (equal to $1/100$) and the log range of the centile, normalized the overall range of growth in the sample.

As a point of comparison, Panel B also shows the fit of a quadratic approximation between the 10th and 90th percentiles of the distribution. In particular, for centile $c \in [10; 90]$, we estimate the following relationship:

$$\log h(g_c) = \alpha - \frac{1}{\Sigma^2}\frac{g_c^2}{2} + \epsilon_c$$

where g_c is the mean growth of centile c and $\log h(g_c)$ the corresponding log density. In the dataset made of these 80 centiles, the R^2 of this regression is 0.98, indicating the fit is quite good in the bulk of the distribution. If the distribution of sales growth was Gaussian, its log-PDF would be well-approximated by a quadratic function for the entire distribution. However, in the top and bottom decile, the log density is much larger than predicted by the quadratic fit, illustrating the presence of non-Gaussian fat tails.

Robustness. In the Appendix, we examine the robustness of our second fact in several ways. First, in Figure A.10, we check the fat tails are not driven by our use of normalized log growth. As expected, we find the opposite is true: in the absence of adjustment, the asymptotic tail coefficient is estimated to be 2, lower than the 2.7 obtained after adjustment. This is consistent with the idea that part of the tail thickness in raw log growth comes from heterogeneous variances. Second, in Figure A.11, we look at percent growth instead of raw growth. We find that percent growth has

thinner tails with an estimated tail coefficnent of 3.5, but is still far from being well-approximated by a Gaussian distribution. Finally, Figure A.12 shows log growth rates adjusted for time variation in mean growth and mean absolute distance in the cross-section. This attempts to correct aggregate changes in mean and volatility of growth. We find that the tails of this distribution are still very thick, with a tail coefficient of 2.5.

2.3 Fact #3: Non-Linear Conditional Expectation

The third and final fact that informs the specification of the data-generating process in our model is that the conditional expectation of current growth is non-linear in past growth. To illustrate this point, Panel A of Figure 5 shows a binned scatter plot of g_{it} versus g_{it-1}. As is evident from the figure, the conditional expectation of g_{it} conditional on g_{it-1} exhibits significant non-linearity reminiscent of the non-linearity in the relationship between forecast errors and revisions in Figure 2.

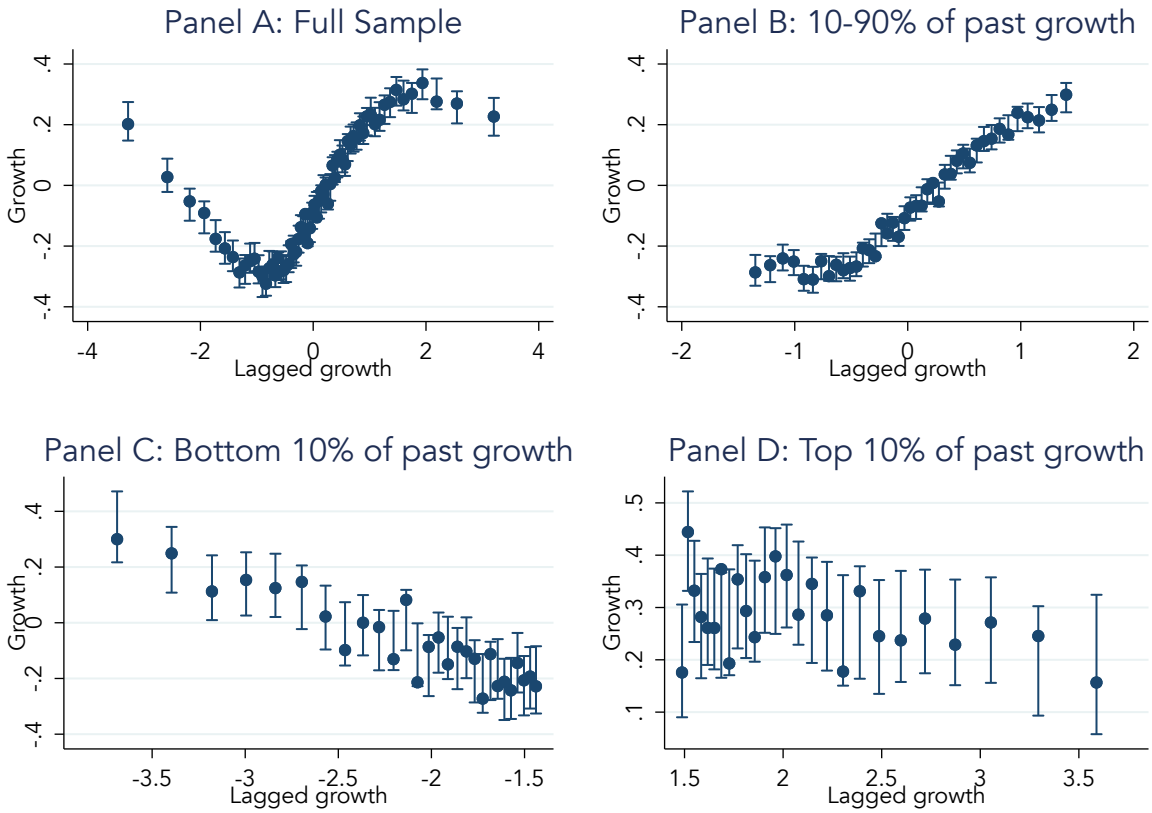
Panels B, C, and D of Figure 5 zoom in on the three parts of the distribution of past growth. These panels illustrate how g_{it} is approximately linearly-increasing in g_{it-1} in the bulk of the distribution, while it is decreasing in g_{it-1} in the tails of the distribution.[4] In Figure A.15, we provide a statistical evidence of non-linearity analagous to Figure 3 by splitting the sample of past growth into 8 quantiles, and then, within each quantile, regressing g_{it} on g_{it-1} (as with revisions we double cluster at the firm and year level). We find the slope is significantly positive in the bulk and significantly negative in the tails of past growth. Put differently, growth appears to be persistent for intermediate levels of past growth, but mean-reverting in the tails. This is consistent with Jaimovich et al. (2025), who find that revenue is more persistent in the bulk of the distribution than in the tails and exhibits fat tails.

3 Model

In this section, we develop a parsimonious model that ties the non-linear relationship between forecast errors and revisions, our Fact #1, to Facts #2 and #3, which are about the data-generating process. We start by describing a model of the DGP that is consistent with our latter two facts, and then turn to a model of belief formation that, given this model of the DGP, generates the first fact.

[4] In Figure A.13 and Figure A.14, we show that this same non-linearity is present for raw log growth, as well as percent growth.

Figure 5: Non-Linear Relationship Between Current and Past Growth



Notes: This figure shows binned scatterplots of current growth on past growth for normalized log sales growth. Panel A shows the results for the entire sample; Panel B restricts the sample the 10-90th percentiles of past growth; Panel C restricts the sample to below the 10th percentile of past growth; Panel D restricts the sample to above the 90th percentiles of past growth.

3.1 Data-Generating Process

The first piece of the model is the data-generating process (DGP) for adjusted log growth, g_t, our forecasting variable of interest. Since this has already been adjusted, our model does not have any firm-specific heterogeneity, and we therefore omit the firm index i for brevity. Without loss of generality, we normalize the unconditional mean of g_t to zero. We then assume that the DGP for g_t takes the following form

$$g_{t+1} = g^*_{t+1} + \sigma_\epsilon \epsilon_{t+1}, \quad \epsilon_t \sim f(\cdot), \tag{2}$$

$$g^*_{t+1} = \rho g^*_t + \sigma_u u_{t+1}, \quad u_t \sim N(0,1), \tag{3}$$

where u_{t+1} and ϵ_{t+1} are IID shocks with a unit variance. The DGP for g_t therefore consists of two components: (i) a persistent component, g_t^*, that follows an AR1 process with normal shocks, and (ii) a transitory shock, ϵ_t, with PDF $f(\cdot)$. Throughout, we assume that g_t^* is an *unobservable* latent state. We denote its unconditional variance by $\sigma_{g^*}^2 = \frac{\sigma_u^2}{1-\rho^2}$, and denote its (normal) PDF by $\phi_{g^*}(\cdot)$. We also denote the marginal PDF of g_t by $h(\cdot)$, which is given by:

$$h(g) = \int_{-\infty}^{+\infty} \phi_{g^*}(g-\epsilon) f(\epsilon) d\epsilon.$$

and denote $\sigma_g^2 = \sigma_\epsilon^2 + \sigma_{g^*}^2$ its unconditional variance.

At this point, if we assumed that ϵ_t was Gaussian, our model of sales growth would be identical to the models of dividend growth in Bansal and Yaron (2004) and Lettau and Wachter (2007). However, unlike these models and existing literature on belief formation, we instead assume the PDF of ϵ_t, $f(\cdot)$, has heavy tails in the sense that it is well-approximated by a power law with tail parameter ν for large values of ϵ_t. Formally, we assume that:

$$f(\epsilon) \propto \epsilon^{-\nu} \text{ as } |\epsilon| \longrightarrow \infty, \quad \nu > 2. \tag{4}$$

Additionally, we assume throughout that $f(\cdot)$ is symmetric.

3.2 Replicating Facts #2 and #3: Fat Tails and Non-Linear $E(g_{t+1}|g_t)$

Before turning to our model of beliefs, we show that our model of the DGP can generate our second and third facts. Our second fact follows from the properties of power laws: the tail parameter of a sum of two independent random variables is the minimum of the tail parameters (Jessen and Mikosch, 2006). Given that g_t is the sum of a normal random variable, g_t^*, and a fat-tailed component, ϵ_t, its tail parameter will be the same as that of ϵ_t, which we denoted by ν.

Showing our model replicates our third fact—the non-linearity in $E(g_{t+1}|g_t)$—is more difficult because there is no closed-form expression for $E(g_{t+1}|g_t)$. Nevertheless, the following proposition leverages a result from empirical Bayes theory known as Tweedie's formula (Efron, 2012) to characterize it.

Proposition 1 (Tweedie's formula)**.** *The expectation of future growth conditional on current growth takes the following form:*

$$E(g_{t+1}|g_t) = -\rho\sigma_{g^*}^2 \frac{d}{dg} \log h(g_t). \tag{5}$$

*In particular, it is a function of the **observable** distribution of growth, g_t.*

Proof. See Appendix B.1. □

Proposition 1 shows that $E(g_{t+1}|g_t)$ is a function of $h(\cdot)$, the marginal PDF of g_t. This result is useful because $h(\cdot)$ is observable, which means we can characterize the shape of $E(g_{t+1}|g_t)$ even without a closed-form. Recall from Panel B of Figure 4 that the growth distribution is approximately Gaussian in the bulk of the distribution: $\log h(g) \approx -\frac{g^2}{2\Sigma^2} + \text{constant}$. Given this approximation, Proposition 1 implies

$$E(g_{t+1}|g_t) \approx \rho \frac{\sigma^2_{g^*}}{\Sigma^2} g_t, \tag{6}$$

which is linear in g_t. This is consistent with Panel B of Figure 5: in the bulk of the distribution, the expectation of g_{t+1} is linearly increasing in g_t.

The intuition for (6) is that, in the bulk of the distribution, g_t is dominated by g^*_t. Since g^*_t is persistent, g_t is then a good approximation of g_{t+1} Note that if g_t were globally Gaussian, then this equation would hold with equality, and we would recover the well-known result that the rational expectations given a signal of a state with additive Gaussian noise is linear. The benefit of Proposition 1 is that it allows us to extend this result to distributions that are locally Gaussian.

Outside of the bulk of the distribution, Panel A of Figure 4 shows that the growth distribution is well-approximated a power law in the tails, consistent with (4). Combining (4) with Proposition 1, we obtain the following corollary.

Corollary 1. *Given the assumption on $f(\cdot)$ in (4), Proposition 1 implies*

$$\lim_{|g_t| \to \infty} E(g_{t+1}|g_t) = \rho \sigma^2_{g^*} \frac{\nu}{g_t}. \tag{7}$$

Proof. See Appendix B.2. □

Corollary 1 shows that, in the tails of growth distribution, the expectation of g_{t+1} is *decreasing* in g_t, unlike in the bulk of the distribution. The intuition for why the conditional expectation is decreasing in g_t is that extreme values of g_t are likely driven by ϵ_t. However, because ϵ_t is not persistent, this value of g_t is unlikely to persist at $t+1$. In the limit as $g_t \to \infty$, g_t reflects ϵ_t with probability one, and $E(g_{t+1}|g_t)$ converges to the unconditional expectation of zero.[5] In sum, (6) and (7) show that our model of the DGP can replicate our third fact.

[5] A related result is developed in Chambers and Healy (2011), who show that when a distribution has sufficiently fat tails such that MLRP does not hold, a positive signal may generate a negative update about the underlying parameter.

Our specification of the DGP is reminiscent of the recent literature on income dynamics (Guvenen et al., 2014, 2021), which models income processes as the sum of persistent and transitory components. The key difference is that fat tails in those models typically come from persistent shocks, unlike our model in which they are transitory. While a model in which u_t had fat tails and ϵ_t was Gaussian could replicate our second fact, Proposition 1 shows that it would be inconsistent with our third fact. Like in our current model, extremely large values of g_t would come from the fat-tailed component of the process with probability one. However, if the fat-tailed component was persistent, this would imply that the conditional expectation of g_{t+1} given g_t would be *increasing* in g_t, which is inconsistent with Figure 5.

3.3 Expectations Formation

Given our model of the DGP, we now describe our model of subjective expectation formation. We assume that expectations do not have "full-information," in the sense that forecasters only observe realizations of g_t but not g_t^*. However, this assumption alone cannot explain our first fact, given it would imply that forecast errors should not be predictable by revisions, which are in forecasters' information set. Therefore, we also assume that forecasts are not rational given this information set. In particular, our core assumption is that forecasters incorrectly perceive distribution of ϵ_t to be Gaussian such that (4) does not hold. We do not micro-found this misperception, but we view it as consistent with the idea that economic agents use simplified, or "sparse", models of reality to form their beliefs (Fuster et al., 2010; Gabaix, 2019). Because of this model misspecification, forecast errors will be (conditionally) predictable.

Since only past values of g_t are observable, agents have to solve a filtering problem to compute their expectations about g_t^* given $g_0, ..., g_t$, which they in turn use to forecast future growth. Under the assumption that agents perceive ϵ_t as being Gaussian, the solution to this filtering problem implies that their expectations will be characterized by the Kalman filter. For our theoretical results, we assume that agents are in a steady-state in the sense that the posterior variance of the Kalman filter and, hence, the Kalman gain is constant. Denoting this steady-state Kalman gain as K, agent's expectations at horizon k, $F_t g_{t+k}$, are characterized by:

$$F_t g_{t+k} = \rho^k K \sum_{s \geq 0} \left(\rho(1-K)\right)^k g_{t-s}. \tag{8}$$

3.4 Replicating Fact #1: Non-Linear Error-Revision Relationship

We now show that our model of belief formation, combined with our model of the DGP that replicates Facts #2 and #3, also generates Fact #1.

Proposition 2. *Define forecast errors and forecast revisions as follows:*

$$ERR_{t+1} = g_{t+1} - F_t g_{t+1},$$
$$REV_t = F_t g_{t+1} - F_{t-1} g_{t+1}.$$

Then, forecast errors are asymptotically linear in forecast revisions:

$$\lim_{|REV_t| \to \infty} E\left(ERR_{t+1} \mid REV_t\right) = -C \times REV_t, \tag{9}$$

where $C > 0$.

Proof. See Appendix B.3. □

Proposition 2 shows that in the tails of the distribution of revisions, the relationship between forecast errors and revisions is linear and negative. This is consistent with our first fact in Figure 2: there is overreaction in the tails of the distribution of revisions. The proof of this result relies on showing that that revisions are driven by changes in g_t, which are asymptotically large (in absolute value) for one of two reasons: (i) a large realization of ϵ_t or (ii) a large realization of past ϵ_{t-h}. The intuition for why these large revisions reflect overreaction is that the forecaster does not realize that they are less likely to be persistent because they are driven by the fat-tailed component of the process, ϵ_t, which is transitory. If ϵ_t was Gaussian, there would still be transitory shocks, but the change in g_t would not be informative about the relative size of transitory and persistent shocks. In contrast, when ϵ_t is fat-tailed, large values of g_t asymptotically only reflect ϵ_t.

Combining Proposition 2 with the fact that the Kalman filter is unbiased on average, we obtain the following result.

Corollary 2. *There exists an* $\overline{R} > 0$ *such that:*

$$E\Big(ERR_{t+1} \times REV_t \,\Big|\, |REV_t| > \overline{R}\Big) < 0$$
$$E\Big(ERR_{t+1} \times REV_t \,\Big|\, |REV_t| < \overline{R}\Big) > 0$$

Proof. See Appendix B.4. □

Corollary 2 shows that our model can generate the full non-linearity in Figure 2. While Proposition 2 shows their is overreaction in the tails, Corollary 2 shows that errors and revisions are positive correlated *on average* in the bulk of the distribution of revisions. The intuition for why forecasters undereact in the bulk is similar to the intuition for overreaction in the tails: the forecasters does not realize that intermediate values of g_t, which generate smaller revisions, are more likely to reflect u_t than ϵ_t, and hence are more likely to be persistent. Note, however, that while Corollary 2 shows that errors and revisions are positively correlated *on average* in the bulk where $|REV_t| < \overline{R}$, we cannot show that this is true for all $|REV_t| < \overline{R}$ without further assumptions of the PDF of ϵ_t, $f(\cdot)$. Additionally, we cannot characterize the exact value of $\overline{R}$ at which the switch between under and overreaction occurs.

3.5 Connection with Augenblick et al. (2024)

We conclude this section by connecting our mechanism for under and overreaction to that in Augenblick et al. (2024). Augenblick et al. (2024) propose a model in which forecasters incorrectly perceive signal quality and shrink it to a default, which leads to overreaction to weak signals and underreaction to strong ones. To illustrate how our model relates to this mechanism, we derive the following result.

Proposition 3. *The variance of future growth conditional on current growth takes the following form:*

$$var(g_{t+1}|g_t) = \sigma_g^2 + \rho^2 \sigma_{g^*}^4 \frac{d^2}{dg^2} \log h(g_t).$$

Proof. See Appendix B.5. □

Proposition 3 shows that the conditional variance of g_{t+1} given g_t depends on the second derivative of the log-density of g_t. Figure 4 shows that $\log h(\cdot)$ is concave in the bulk (Panel B), but convex in the tails (Panel A). Using Proposition 3, these patterns imply that g_t is a more precise signal of g_{t+1} for smaller values of past growth, while it is less precise in the tails of the distribution. Therefore, our model delivers a similar result to Augenblick et al. (2024): forecasters are underreacting in the parts of the distribution where signals are strong, and overreacting when signals are weak.

4 Quantitative Fit of the Model

This section assesses whether the model in Section 3 can quantitatively account for our three main empirical facts.

4.1 Simulation Details

The model that we take to the data is the same model described in Section 3 with two changes. First, we assume that ϵ_t is distributed according to a t-distribution with $\nu > 2$ degrees of freedom normalized to have a unit variance. The t-distribution is asymptotically power law with tail parameter ν, and has the nice property of converging to a normal distribution as $\nu \longrightarrow \infty$. Second, we relax the assumption that the Kalman filter updating equations are applied using a constant Kalman gain, which would only apply in a steady state. Instead, we use the following updating equations to compute subjective forecasts, which follow from applying standard Kalman filter results to equations (2) and (3) under the (incorrect) assumption that $\epsilon_t \sim N(0,1)$:

$$\begin{aligned} F_t g_{t+h} &= \rho^h F_t g_t^* \\ F_t g_t^* &= (1-K_t) F_{t-1} g_t^* + K_t g_t \\ K_t &= \frac{\Sigma_t}{\Sigma_t + \sigma_\epsilon^2} \\ \Sigma_{t+1} &= \rho^2 (1-K_t) \Sigma_t + \sigma_u^2 \\ F_0 g_0^* &= g_0^*, \quad \Sigma_0 = 0 \end{aligned} \tag{10}$$

In our simulations, we sample time series of g_t according to equations (2) and (3) with 100,000 observations. We repeat this simulation 100 times, where g_0^* is drawn from its stationary distribution, and the length of the simulation burn-in period is 50 observations for each series. This gives us a total of 10 million simulated observations.[6]

[6]We choose this simulation size to be as large as possible without exceeding the RAM of our GPU. The simulation itself is not memory or computationally intensive—these constraints only become binding because when we compute the rational expectation using a particle filter described below.

4.2 Estimating Parameters of the DGP

We start by estimating the parameters of the DGP g_t (equations (2)-(3)). Our model has four parameters that control the data-generating process: ρ, the persistence of g_t^*, σ_ϵ, the scale parameter for ϵ_t, ν, the tail parameter of ϵ, and the innovation volatility, σ_u. We estimate these parameters using Simulated Minimum Distance (SMD), minimizing the difference between a set of statistics computed from simulated data in the model and the corresponding value of those statistics in the data. We use the inverse-diagonal covariance as the weighting matrix.[7]

We use 9 moments to estimate the 4 parameters. First, we use 6 slope coefficients from regressing g_{t+1} on g_t for 6 different subsamples based on percentile breakpoints of g_t: $[0; 10]$, $[10; 20]$, $[20; 50]$, $[50; 80]$, $[80; 90]$, $[90; 100]$. We also ask the estimated model to match four additional moments: (1) the Pareto tail coefficient from Panel A of Figure 4, (2) the 10-90 percentile difference of g_t (3) the 10-90 percentile difference of $g_{t+1} - g_t$, and (4) the 10-90 percentile difference of $g_{t+3} - g_t$.

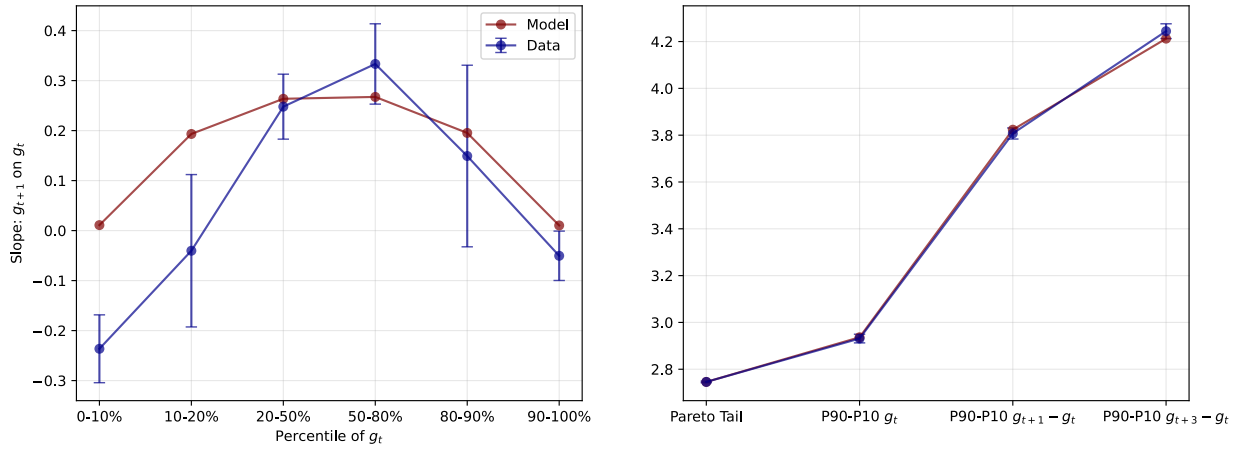
The idea behind these moments is to try to get the model to match Facts #2 and #3 in the data. The first six regression slopes capture Fact #3: the non-linearity in the conditional expectation of g_{t+1} given g_t in Figure 5.[8] These regression coefficients jointly identify ρ, σ_u, and ν, but they do not separately identify each of these parameters. Therefore, we include a seventh statistic that captures Fact #2: the Pareto tail coefficient in Figure 4. This moment is useful because it is only affected by ν and, hence, can be used to separately identify it. To summarize, the 6 slopes and the tail coefficient jointly help pin down the Gaussian and non-Gaussian part of the process, but not its scale.

The eighth statistic, the dispersion of g_t, serves this purpose. We choose to use the spread between the 10th and 90th percentiles to avoid sensitivity to outliers, which inevitably occur in a process with fat tails. This statistic primarily identifies the scale of the DGP, in particular σ_u and σ_ϵ. The final two statistics that we use are the spread between the 10th and 90th percentiles of one- and three-year changes in g_t. These additional statistics are useful for separately identifying ρ and σ_u. Without these moments, the estimation has several local minima with different values of ρ and σ_u. Including these moments is useful because the extent to which the three-year changes tend to larger than one-year changes is directly affected by ρ and no other parameters. Our choice of one- and three-year changes, specifically, follows the literature on income dynamics, which uses the same

[7] Because the SMD objective function depends on four parameters, we need to be careful that we reach a global rather than local minimum. We perform this optimization using a two-step procedure in which we first search of a quasi-random grid of 20,000 points, and then run local Nelder-Mead optimizations using the top 10 points as starting points. Our result is then the parameter vector that has the lowest objective function from any of these Nelder-Mead optimizations. We have verified that with our choice of statistics, these local optimizations all converge to similar points.

[8] Before running these regressions in both the model and the data, consistent with our empirical analysis, we trim observations at the 1st and 99th percentiles to avoid the influence of extreme outliers.

Figure 6: Fit of Estimated Model on Data-Generating Process



Notes: This figure shows the fit of the estimated model on the statistics used to estimated it. The values of the statistics in the data are shown with 95% confidence intervals. The values in the model are computed from simulations, as described in Section 4.1, at the set of parameters shown in (11).

moments to identify persistence and volatility in similar data-generating processes (e.g., Guvenen et al., 2014, 2021).

Our SMD procedure delivers the following estimates:

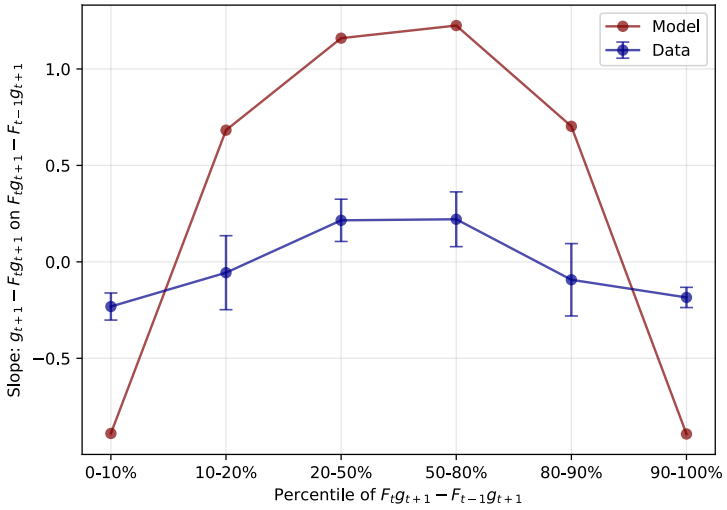
$$\rho = 0.529, \sigma_u = 0.631, \sigma_\epsilon = 1.325, \nu = 2.533. \tag{11}$$

Figure 6 shows the fit of the model on the targeted statistics using the parameters in (11).The results show that the model is able to fit all the moments in the data relatively well. The main dimension on which the model misses is relationship between g_{t+1} and g_t in the left tail of the distribution of g_t. In the data, there is some mean reversion that manifests in a negative slope coefficient, which the model cannot generate.

4.3 Estimating the Expectations Formation Model

We now turn to the model fit of our first fact. First, we note that the expectations model (10), based on the Kalman filter, generates "too much" bias on our estimated DGP. We show this in Figure 7. Consistent with our theoretical results in Section 3.4, the Kalman filter generates a non-linear relationship between errors and revisions, with positive coefficients (i.e., underreaction) for intermediate revisions and negative coefficients (i.e., overreaction) for large revisions. However, relative to the data, the model generates too much non-linearity: the error-revision slope is too positive in the bulk of revisions, and too negative in the tails. The fact that our model cannot fit the

Figure 7: Fit of Estimated Model on Fact #1: Non-Linearity in Error-Revision Relationship



Notes: This figure shows the fit of the model on the error-revision relationship in the data. The values of the statistics in the data are shown with 95% confidence intervals, which are repeated from Figure 3. The values in the model are computed from simulations, as described in Section 4.1, at the set of parameters shown in (11). The left panel shows results when agents form forecasts using a simplified short-memory linear rule based on the population regression coefficient of g_{t+1} on g_t. The right panel shows results when agents form forecasts according to the Kalman filter in (10). Before running these regressions in both the model and the data, we trim observations at the 1st and 99th percentiles to avoid the influence of extreme outliers.

data is not surprising given that our model of beliefs has no additional degrees of freedom, given our estimates of the DGP parameters.

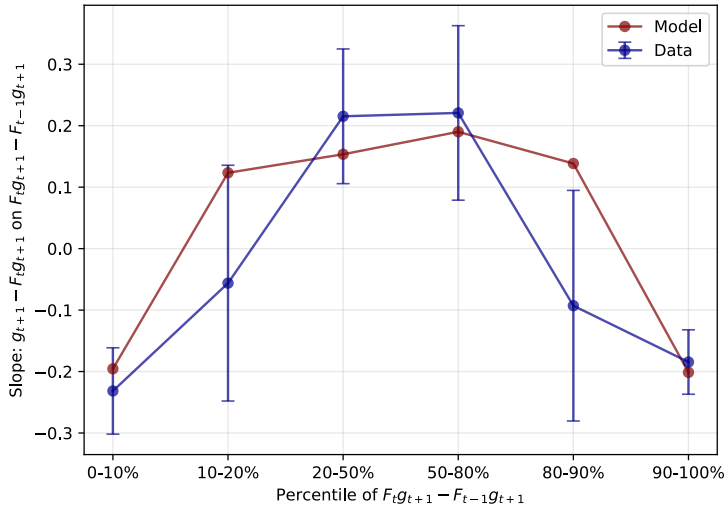
To attenuate the bias of our linear forecasting model, we allow forecasts to be anchored in rational expectation (as in Fuster et al. 2010 and Gabaix 2019). In particular, we assume that expectations are a weighted-average of the Kalman filter and the true (non-linear) rational expectation:

$$F_t^{\lambda} g_{t+h} = \lambda F_t g_{t+h} + (1 - \lambda) E_t g_{t+h} \tag{12}$$

where E_t is the rational expectation given all past history of g_t and λ is to the weight that is placed on Kalman filter forecast. $\lambda = 1$ corresponds to the case in Figure 7, while $\lambda = 0$ corresponds to the case of rational expectations in which error-revision coefficients would always be zero.

We estimate λ to assess that the anchored model in (12) can match our first fact quantitatively. In order to do this, we first need to compute the rational expectation, $E_t g_{t+1}$, which we do using the particle filtering algorithm from Fernandez-Villaverde and Rubio-Ramirez (2007) (also known as sequential importance sampling). See Appendix C for a detailed description of this procedure. Having computed the rational expectation, we then estimate λ using SMD with six statistics and the inverse-diagonal covariance as the weighting matrix. The six statistics that we use are the

Figure 8: Fit of Estimated Model with $\lambda = 0.290$ on Fact #1: Non-Linearity in Error-Revision Relationship



Notes: This figure shows the fit of the estimated model with the forecasting equation (12). The values of the statistics in the data are shown with 95% confidence intervals. The values in the model are computed from simulations, as described in Section 4.1, at the set of parameters shown in (11) and (13).

error-revision regression coefficients in Figure 7. This SMD procedure delivers the following estimate:

$$\lambda = 0.290. \tag{13}$$

Our estimated value of λ implies that forecasters place 29% of their weight on the Kalman filter forecast, and 71% weight on the true forecast. Figure 8 shows the fit of the model on the targeted error-revision coefficients using the parameters in (11) and (13). The fit is much better than in Figure 7, where the model predicts way too much non-linearity. Now almost all of the error-revision coefficients are within the 95% confidence intervals of the data. One area where the model misses is that it generates positive rather negative coefficients in the 10–20% and 80–90% bins, like in the case of the Kalman filter in Figure 7.

4.4 Accuracy Loss Relative to Rational Expectation

Having computed the (limited-information) rational expectation, we next examine its accuracy relative to the Kalman filter and the forecasts formed in (12) with the estimated value of λ. Table 2 shows the percent loss in mean-squared error (MSE) of these two forecasts in separate panels relative to the rational expectation. The different rows correspond to different values of ρ, while the different columns correspond to different values of ν. At the estimated values of $\rho = 0.529$

and $\nu = 2.533$, the loss in MSE from using the Kalman filter is around 1.2%, which is non-trivial. This loss tends to increase as ρ increases and ν increases, which is to be expected. At low values of ρ, forecasts of g_{t+1} are not very dependent on the solution to the filtering problem, since g_t^* is not persistent. At high values of ν, the $f(\cdot)$ becomes closer to a normal distribution, meaning the Kalman filter provides a closer approximation to the true solution. In contrast, the loss in MSE from agents forecasts with $\lambda = 0.290$ is much smaller. At the estimated values of ρ and ν, the loss is around 0.1%. Like in Panel A, this loss increases as ρ increases and ν decreases. Overall, the cost of forecasting errors induced by the Kalman filter is modest, which supports the idea that it would be a "default" in a model of bounded rationality (Fuster et al., 2010; Gabaix, 2019).

An important caveat to the analyses in Table 2 is that both the rational expectation, which we compute using the particle filter, and the Kalman filter assume that forecasters know the parameters of the DGP. If forecasters instead have to learn these parameters, as in Farmer et al. (2024) and Singleton (2021), then it is possible that the optimal forecast may ignore the presence of fat tails if they are sufficiently difficult to detect in finite samples.

5 Additional Tests of Model Predictions

This section presents two additional tests of our model's predictions. The first one comes from an online forecasting experiment, and the second comes from data on stock returns.

5.1 Evidence from a Forecasting Experiment

5.1.1 Experimental Design

The design of our online forecasting experiment is taken from Afrouzi et al. (2023) (AKLMT). Participants are asked to predict the outcome of a process, and their compensation depends on the accuracy of their forecasts. We recruit participants on Amazon MTurk, and they are reasonably representative of the general population. The experiment does not require participants to have any prior knowledge of statistics and participants do not know the DGP, although AKLMT show that this is not important. The interface is graphical and user-friendly: participants click with their mouses to provide their forecasts at one- and two-period ahead horizons. Prior to making their first forecast, they see 40 prior realizations of the process, and then sequentially provide both forecasts in 40 periods, seeing the realization of the process between each period. We refer the reader to AKLMT

Table 2: Percent Loss in One-Period Ahead Mean-Squared Error Relative to Rational Expectation

Panel A: Kalman Filter Forecast

				ν				
ρ	2.1	2.5	2.533	3.0	3.5	4.0	4.5	5.0
0.1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.2	0.1%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.0%
0.3	0.2%	0.4%	0.4%	0.3%	0.2%	0.2%	0.1%	0.1%
0.4	0.4%	0.7%	0.7%	0.6%	0.4%	0.3%	0.2%	0.2%
0.5	0.6%	1.1%	1.1%	0.9%	0.7%	0.5%	0.4%	0.3%
0.529	0.7%	1.2%	1.2%	1.0%	0.8%	0.6%	0.4%	0.3%
0.6	1.0%	1.6%	1.6%	1.3%	1.0%	0.7%	0.5%	0.4%
0.7	1.4%	2.2%	2.2%	1.8%	1.4%	1.0%	0.8%	0.6%
0.8	1.9%	3.0%	2.9%	2.4%	1.8%	1.4%	1.1%	0.8%
0.9	2.5%	3.6%	3.6%	3.0%	2.3%	1.8%	1.4%	1.1%

Panel B: Forecast in (12) with Estimated λ

				ν				
ρ	2.1	2.5	2.533	3.0	3.5	4.0	4.5	5.0
0.1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.4	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
0.5	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
0.529	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
0.6	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%
0.7	0.1%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.0%
0.8	0.2%	0.3%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
0.9	0.2%	0.3%	0.3%	0.2%	0.2%	0.1%	0.1%	0.1%

Notes: This table shows the percent loss in mean-squared error (MSE) in model simulation of different one-period ahead forecasts relative to the one-period ahead rational expectation computed using Algorithm 1. Panel A shows the results for the Kalman filter forecast, computed as in (10); Panel B shows the results for the forecast in (12) with the estimated value of λ in (13). The different rows correspond to different values of ρ, while the different columns correspond to different values of ν. For each different combination of ρ and ν, σ_ϵ and σ_u are adjusted so that the analytical variance of g_t and g_t^* remain the same as in the model with the estimated parameters in (11).

for further details about this design; see Figure A.16 an example of the interface our participants see for these parameters.

Starting with this design, we set the DGP of the process being forecasted to the following:

$$g_{t+1} = g^*_{t+1} + 12.16\epsilon_{t+1}, \quad \epsilon_t \sim t(2.533) \tag{14}$$

$$g^*_{t+1} = 0.529 g^*_t + 12.62 u_{t+1}, \quad u_t \sim N(0,1) \tag{15}$$

This DGP corresponds to the one in our model with the estimated values of ρ and ν from Section 4. We have modified σ_u and σ_e, which simply scale the DGP, so that the values are larger numbers that are easier for participants to interpret. We then ran the experiment in two waves with 201 participants on March 17th, 2025 and 202 participants on April 29, 2025. Given that each participant makes 40 forecasts, our data has 16,120 observations. We compare our results with data from a condition in AKLMT in which the process follows an AR1 process with Gaussian shocks and a persistence parameter of 0.2.[9]

Before describing the results, we briefly discuss the relative benefits and costs of this experiment relative to our evidence in Section 2. The main benefit of running an experiment is that it allows us to directly control the parameters of the DGP. However, this benefit comes with several downsides. First, participants in the experiment are fundamentally different from the forecasters in our data, who are professional equity analysts with incentives aside from accuracy. Second, the professional forecasters in our data are likely to have a better understanding of the DGP, while we cannot control the priors of the participants in our experiment. Finally, the environment of forecasting a variable in our experiment is obviously quite different from forecasting in the real-world. Despite these differences, we still view this experiment as a useful because, by comparing our results with those in AKLMT, we can perform a direct test of our theory's key prediction that transitory fat-tailed shocks create a non-linear relationship between errors and revisions.

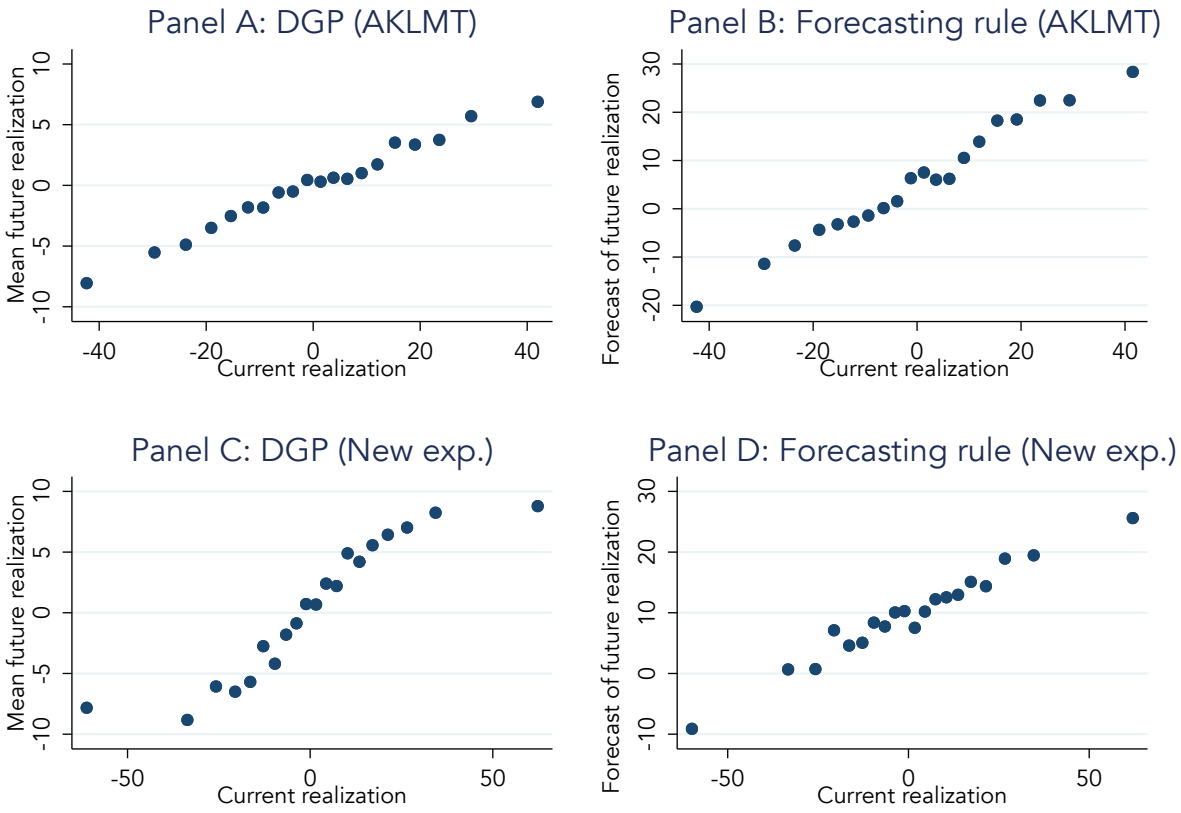
5.1.2 Experiment Results

Panel A of Figure 9 shows the relationship between g_{it+1} and g_{it} in the data from AKLMT, where i is the participant and t the round of forecasting. As expected, this relationship is linear. In contrast, Panel C shows the same relationship in our experimental data, where the DGP is in (14) and (15). Although our experimental data has an order of magnitude less observations than our data on sales growth, these data still replicate this non-linear relationship in Fact #3.

Panels B and D of Figure 9 provide evidence that, despite the differences in the DGP, participants' one-period ahead forecasts, $F_{it}g_{it+1}$, are still linear in g_{it}, unlike the true DGP. This is consistent with our conjecture the forecasters ignore the fact that large shocks are likely to be transitory. Comparing Panels A and C with B and D also illustrates that forecasters substantially overestimate the persistence of the process: $E\left(F_{it}g_{it+1} \mid g_{it}\right)$ is steeper than $E\left(g_{it+1} \mid g_{it}\right)$ at all points. The combination of this extrapolation and the likely presence of expectation noise (de Silva and Thesmar,

[9]We use 0.2 because it is closest experimental condition in AKLMT to the regression coefficient of g_{it+1} and g_{it} in our model estimated in Section 4.

Figure 9: Data-Generating Process and Forecasts in Experimental Data



Notes: Panels A and C show binned scatterplots of g_{it+1} versus g_{it} in experimental data, where i is the participant and t the round of forecasting. Panels B and D show binned scatter plots of $F_{it}g_{it+1}$ against g_{it}. Panels A and B use data from Afrouzi et al. (2023), where the DGP is a Gaussian AR1 process with a persistence parameter of 0.2. Panels C and D use the experimental data described in the main text, where the DGP is fitted on our data – see equations (14) and (15).

2024) leads to significant overreaction on average: columns (1) and columns (3) of Table 3 show that the relationship between forecast errors is significantly negative in our experiment and in AKLMT, with a similar magnitude. Given both processes are not very persistent, this finding is consistent with the conclusion in AKLMT that that overreaction is especially strong for processes that are less persistent.

Despite the presence of overreaction on average, the fact that $E\left(F_{it}g_{it+1} \mid g_{it}\right)$ is linear in Panel D while $E\left(g_{it+1} \mid g_{it}\right)$ is non-linear in Panel C provides preliminary evidence that forecasters are likely to overreact more in the tails of the distribution, consistent with our theory. To test for this

non-linearity, we estimate the following regression on the data in AKLMT and from our experiment:

$$ERR_{it+1} = \alpha + \beta REV_{it} + \gamma_L \text{Bottom 40\% } REV_{it} + \gamma_H \text{Bottom 40\% } REV_{it} + \beta_L REV_{it} \times \text{Bottom 40\% } REV_{it} + \beta_H REV_{it} \times \text{Top 40\% } REV_{it} + e_{it+1} \quad (16)$$

where Bottom 40% REV_{it} and Top 40% REV_{it} are indicator variables that equal one when revisions are in the bottom or top 40% of the distribution. Our prediction is that the relationship between errors and revisions is more negative in the tails, so β_H and β_L should be negative and of similar magnitudes in our data and not different from zero in the AKLMT data. We choose to split the data into 0-40%, 40-60%, and 60-100% based visually inspecting the relationship forecast errors and revisions in our data. The choice of larger groups in the tail of the distribution reflect the fact that forecasters overreact on average, as described above, in both AKLMT and our data. Nevertheless, our qualitative conclusions are robust to alternative splits of the data.

Columns (2) and (4) of Table 3 shows that the results from estimating (16) are consistent with our theory. For the data from AKLMT, column (2) shows no difference in the amount of overreaction that is present in the bulk of the distribution and the tails: β_L and β_H are not statistically different from zero. In contrast, in our data, column (4) shows that β_L and β_H are significantly negative with similar magnitudes, while β is significantly positive. Collectively, these findings provide direct evidence for our theory's key prediction that transitory fat-tailed shocks create a non-linear relationship between errors and revisions.

5.2 Implications for Return Predictability

This section examines and tests the predictions that our model of belief formation makes for return predictability. We focus on one type of predictability: momentum, the fact that past returns predict future returns (Jegadeesh and Titman, 2011).

5.2.1 Model Prediction: Non-Linearity in Momentum

To derive predictions about returns, we two simplifying assumptions: (i) earnings growth is a constant fraction of sales growth, and (ii) the subjective discount rate is constant (as in Bouchaud et al. 2019 and Nagel and Xu 2019), which is consistent with evidence in De la O and Myers (2021). These assumptions allow us to simulate a dataset of realized returns consistent with realized and expected earnings generated by our estimated DGP and forecasting model. As shown in

Table 3: Non-Linear Relationship Between Forecast Errors and Revisions in Experimental Data

	Dependent Variable: Error			
	Afrouzi et al. (2023)		New Experiment	
	(1)	(2)	(3)	(4)
Revision	-0.44***	-0.39	-0.41***	0.97**
	(0.02)	(0.41)	(0.01)	(0.39)
Revision × Bottom 40 %		-0.03		-1.46***
		(0.42)		(0.39)
Revision × Top 40 %		-0.11		-1.41***
		(0.41)		(0.40)
Bottom 40 %		-4.49		-2.84
		(2.95)		(2.48)
Top 40 %		-2.53		5.79**
		(2.73)		(2.36)
Constant	-6.62***	-2.84	-8.61***	-10.71***
	(1.46)	(2.50)	(1.33)	(2.35)
Number of Observations	6,942	6,942	15,717	15,717
Gaussian AR1 Process	✓	✓	✗	✗
Fat-Tailed Process	✗	✗	✓	✓

Notes: This table shows results from regressions of forecasts errors onto revisions In columns (1) and (2), we use data from Afrouzi et al. (2023), where the DGP is a Gaussian AR1 process. In columns (3) and (4), we use the experimental data described in the main text, where the DGP is fitted on our data – see equations (14) and (15). Columns (1) and (3) just regress errors on revision. Columns (2) and (4) estimate equation (16). All standard errors are clustered at the participant level. These regressions have fewer than 16,120 observations because computing forecasts revisions loses one observation per subject.

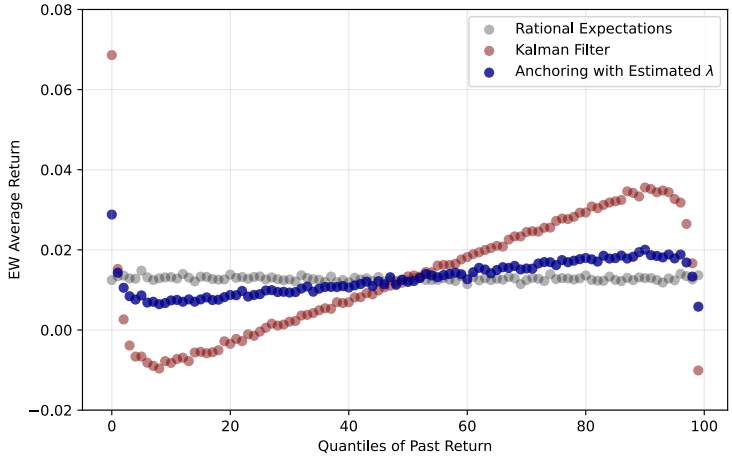
Appendix B.6, the Campbell (1991) decomposition coupled with these assumptions allows us to link returns with changes in expectations of sales growth g_t using the following expression:

$$\log(1+R_t) = \log(1+R_f+\pi) + (g_t - F_{t-1}g_t) + \sum_{k=1}^{\infty} c^k REV_t g_{t+k}, \tag{17}$$

where R_t is the net return, π is the (constant) ERP, R_f is the gross risk-free rate, c is a linearization constant, and $REV_t g_{t+k}$ denotes investors' subjective revisions about future growth between $t-1$ and t. This expression is intuitive: given that discount rates are fixed, returns are driven earnings surprises and revisions of future growth. As detailed in Appendix B.6, we use (17) to generate a panel of simulated returns based on panel of realized and expected sales growth. The parameters used are the ones estimated in the previous section.

Figure 10 shows the prediction that our model makes for momentum by showing a binned scatterplot of (log) returns r_t on past returns r_{t-1}. The model predicts momentum in the bulk of past

Figure 10: Model-Implied Relationship Between Current and Future Returns



Notes: This future shows a binned scatterplot of future returns against current returns in our simulated model. Given a set of earnings growth expectations, we compute returns using (17), as described in Appendix B.6. The three sets of points on the graph correspond to three cases in which beliefs are set equal to (i) the rational expectation computed with the particle filter, (ii) the Kalman filter, and (iii) the combination of the former two as in (12) with the value of λ in (13). We conduct this simulation assuming $R_f = 1.01$, $\pi = 5.5\%$, $c = 0.96$, and ρ is set to the estimated value in (11). We set the constant of proportionality between sales and earnings growth such that the standard deviation of returns is 15%.

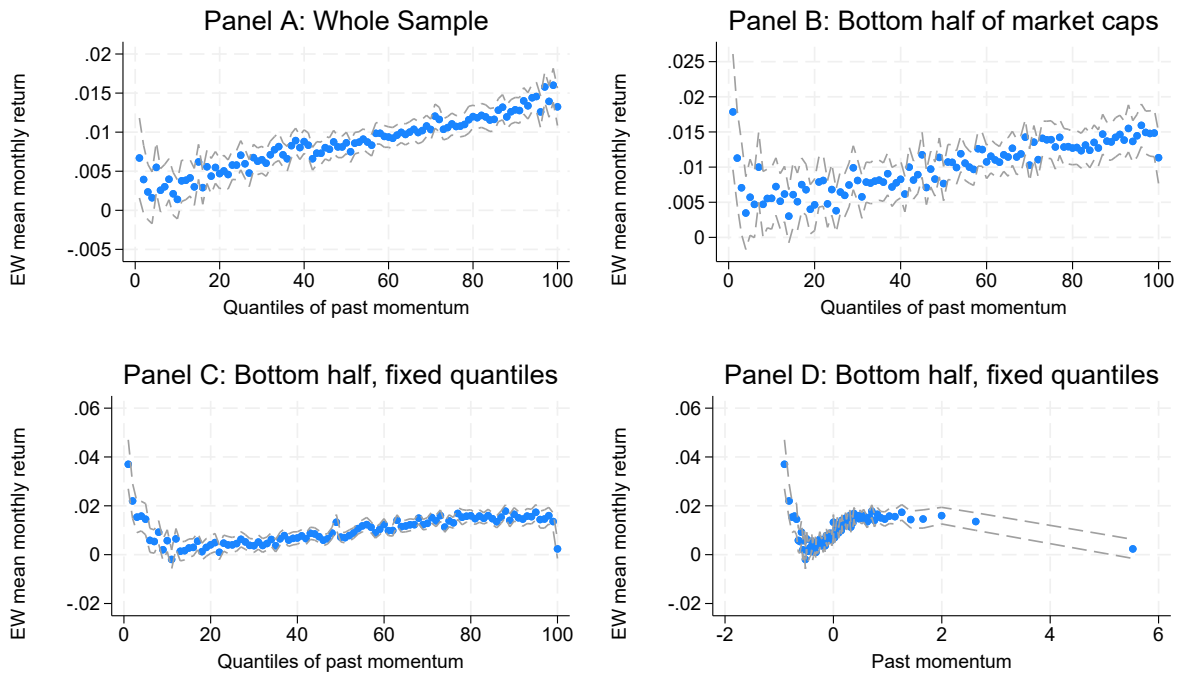
returns, but mean-reversion in the tails. This is consistent with our expectations formation model which predicts overreaction to large news, and underreaction to intermediate ones. As the Figure also shows, predictability is stronger for the more "biased" expectation: Kalman filter expectations (blue dots) generate the strongest predictability, while forecasts anchored to rational expectations (red dots) generate weaker predictability. Predictability disappears completely when forecasts are rational (gray dots) or when ϵ is Gaussian, in which case the Kalman filter is rational.

5.2.2 Momentum Non-Linearity in the Data

We now examine if momentum indeed exhibits mean-reversion in the tails in the data. We use CRSP monthly returns from 1927 to 2023, restrict ourselves to all firms listed on NYSE, AMEX and NASDAQ, and adjust returns for delisting. Our measure of momentum follows the literature (Jegadeesh and Titman, 2011): every month t, momentum is the cumulative return between months $t-12$ and $t-1$ (thereby excluding the last month of past returns, which exhibit reverals). We also compute firm size as the market capitalization 12 months prior to t.

Motivated by Figure 10, we first plot returns against ceniles of momentum that are redefined each month, as in standard asset pricing tests. Panel A shows the results for the entire CRSP sample: the line is upward sloping, consistent with the presence of momentum. However, we also find that

Figure 11: Relationship Between Current and Past Returns in the Data



Notes: This figure shows binned scatterplots of monthly returns, r_t, as function of past annual returns excluding month $t-1$, $r_{t-12,t-1}$. There are 100 bins in each panel. Panels A-C use the standard convention of using bins of past returns $r_{t-12,t-1}$ as the x-axis. Panel D uses past returns, $r_{t-12,t-1}$, directly as the x-axis. Panel A uses the entire CRSP sample of stocks traded on AMEX, NASDAQ and NYSE. Panel B uses only the bottom half of stocks ranked by 12 months lagged market cap. In contrast to Panels A and B, quantiles of returns are defined on the entire sample in Panels C and D, rather than separately for each month. Dashed lines are 95% confidence bands based on standard errors assuming that returns are independent.

there is a bit of mean-reversion for "super losers", but not for "extreme winners". Panel B shows that the non-linearity is a bit more pronounced for smaller firms, which are defined as firms with below-median size in each month. This is to be expected as smaller stocks are more expensive to trade and therefore should display more predictability (Novy-Marx and Velikov, 2016). In Panels C and D, we define the centiles on the entire sample, thereby accounting for times series variation in momentum returns. Panel C shows that, using fixed centiles and focusing on firms below the 50% size cutoff, the non-linear relationship starts appearing significantly. This suggests that the time series of momentum returns generate some mean-reversion in the tails: when volatility is high, losers are more likely to overperform, and winners more likely to underperform. This is consistent with the literature on momentum crashes (Clara and Barroso, 2015) and with our model, to the extent that tail risk emerges when aggregate volatility is high. The quantitative shape of this relationship in Panel C is also remarkably similar to that in Figure 10. Panel D reproduces the analysis of Panel C, except that the x-axis is now rescaled to the average values of past returns. This

last graph is less consistent with practice of forming portfolios, as is typically done in asset pricing tests. Again, there is significant non-linearity in the return-momentum relationship, in line with the prediction of our model.

6 Conclusion

In this paper, argue that recognizing the complexity of the underlying DGP, in particular it's fat tails are non-Gaussian dynamics, are crucial for understanding the properties of subjective forecasts. We document three facts using data on sales growth forecasts by equity analysts: (i) the relationship between forecast revisions and future forecast errors—the variables used in Coibion and Gorodnichenko (2015) regressions—is strongly non-linear; (ii) the distribution of the underlying process has fat tails; and (iii) the conditional expectation of future sales growth is non-linear in current growth, with mean reversion in the tails. Next, we build a forecasting model that connects these facts. The key ingredients in our model are that the underlying process is non-Gaussian, but forecasters fail to recognize this. After showing formally that our model can explain the three facts we documented in the data, we estimate it and show that it does so quantitatively. Finally, we show that our framework is consistent with evidence from an online forecasting experiment where the underlying process is non-Gaussian and that it provides an explanation for non-linearity in the momentum of stock returns.

Our paper raises several questions for further work. First, our model of belief formation is reduced-form. It would be fruitful to try to provide a micro-foundation for why forecasters ignore fat tails in data-generating processes, which would allow us to study how this bias would manifest for other data-generating processes. Second, it would be useful to try to estimate our shrinkage parameter using other data on subjective forecasts. Variation in this shrinkage parameter across different data-generating processes would be useful for disciplining a more micro-founded model of belief formation. Finally, our paper raises the question of how learning occurs in an environment with fat-tailed processes, which is likely to occur much more slowly and be more difficult.

References

Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar, 2023, Overreaction in expectations: Evidence and theory, *Quarterly Journal of Economics* .

Augenblick, Ned, Eben Lazarus, and Michael Thaler, 2024, Overinference from weak signals and underinference from strong signals, *Quarterly Journal of Economics* .

Axtell, Robert L., 2001, Zipf Distribution of U.S. Firm Sizes, *Science* 293, 1818–1820.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The Journal of Finance* 59, 1481–1509.

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma, 2022, Belief distortions and macroeconomic fluctuations, *American Economic Review* 112, 2269–2315.

Boar, Corina, Denis Gorea, and Virgiliu Midrigan, 2025, Why are returns to private business wealth so dispersed?, *Working Paper* .

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020a, Overreaction in Macroeconomic Expectations, *American Economic Review* 110, 2748–82.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2019, Diagnostic Expectations and Stock Returns, *The Journal of Finance* 74, 2839–2874.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2020b, Memory, Attention, and Choice*, *The Quarterly Journal of Economics* 135, 1399–1442.

Bottazzi, Giulio, and Angelo Secchi, 2006, Explaining the distribution of firm growth rates, *The RAND Journal of Economics* 37, 235–256.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar, 2019, Sticky Expectations and the Profitability Anomaly, *Journal of Finance* 74, 639–674, Publisher: Wiley Online Library.

Campbell, John Y., 1991, A Variance Decomposition for Stock Returns, *The Economic Journal* 101, 157, Publisher: Oxford University Press (OUP).

Campbell, John Y., and Robert J. Shiller, 1988, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195–228.

Chambers, Christopher P., and Paul J. Healy, 2011, Reversals of signal-posterior monotonicity for any bounded prior, *Mathematical Social Sciences* 61, 178–180.

Clara, Pedro Santa, and Pedro Barroso, 2015, Momentum has its moments, *Journal of Financial Economics* 116, 111–120.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts, *American Economic Review* 105, 2644–2678.

De la O, Ricardo, and Sean Myers, 2021, Subjective Cash Flow and Discount Rate Expectations, *The Journal of Finance* 76, 1339–1387.

de Silva, Tim, and David Thesmar, 2024, Noise in expectations: Evidence from analyst forecasts, *Review of Financial Studies* .

Denuit, Michel, Patricia Ortega-Jimenez, and Christian Robert, 2024, Conditional expectations given the sum of independent random variables with regularly varying densities, Technical report, LIDAM Discussion Paper ISBA.

Efron, Bradley, 2012, Tweedie's formula and selection bias, *Journal of the American Statistical Association* .

Farmer, Leland E., Emi Nakamura, and Jón Steinsson, 2024, Learning About the Long Run, *Journal of Political Economy* 132, 3334–3378.

Fernandez-Villaverde, Jesus, and Juan F. Rubio-Ramirez, 2007, Estimating macroeconomic models: A likelihood approach, *Revuew of Economic Studies* 74, 1059–1087.

Fuster, Andreas, David Laibson, and Brock Mendel, 2010, Natural Expectations and Macroeconomic Fluctuations, *Journal of Economic Perspectives* 24, 67–84, Publisher: American Economic Association.

Gabaix, Xavier, 2009, Power Laws in Economics and Finance, *Annual Review of Economics* 1, 255–294.

Gabaix, Xavier, 2019, Behavioral Inattention, *Handbook of Behavioral Economics* .

Graeber, Thomas, Christopher Roth, and Marco Sammon, 2024, Categorical processing in an imperfect world, Technical report.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song, 2021, What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?, *Econometrica* 89, 2303–2339.

Guvenen, Fatih, Serdar Ozkan, and Jae Song, 2014, The nature of countercyclical income risk, *Journal of Political Economy* 122, 621–660.

Jaimovich, Nir, Stephen J. Terry, and Nicholas Vincent, 2025, The empirical distribution of firm dynamics and its macro implications, *Working Paper* .

Jegadeesh, Narasimhan, and Sheridan Titman, 2011, Momentum, *Annual Review of Financial Economics* 3, 493–509.

Jessen, Anders Hedegaard, and Thomas Mikosch, 2006, Regularly varying functions, *Publications de l'institut mathématique* 79, 2303–2339.

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran, 2020, The Tail that Wags the Economy: Belief-Driven Business Cycles and Persistent Stagnation, *Journal of Political Economy* 128, 2839–2880.

Kwon, Spencer, and Johnny Tang, 2025, Extreme categories and overreaction to news, *Review of Economic Studies, forthcoming* .

Lettau, Martin, and Jessica A. Wachter, 2007, Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium, *The Journal of Finance* 62, 55–92.

Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar, 2020, A Quantitative Analysis of Distortions in Managerial Forecasts, Technical Report w26830, National Bureau of Economic Research, Cambridge, MA.

Moran, José, Angelo Secchi, and Jean-Philippe Bouchaud, 2024, Revisiting granular models of firm growth.

Nagel, Stefan, and Zhengyang Xu, 2019, Asset Pricing With Fading Memory, *Review of Financial Studies* Forthcoming.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Robbins, Herbert, 1956, An empirical bayes approach to statistics, *Berkeley Symposium on Mathematical Statistics and Probablility* 157—163.

Singleton, Kenneth J., 2021, Presidential address: How much "rationality" is there in bond-market risk premiums?, *The Journal of Finance* 76, 1611–1654.

Stanley, Michael, Luis Amaral, Sergey Bouldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael Salinger, and Eugene Stanley, 1996, Scaling behavior in the growth of companies, *Nature* 379, 804–806.

Wang, Chen, 2021, Under- and overreaction in yield curve expectations, *Working Paper* .

Wyart, Matthieu, and Jean-Philippe Bouchaud, 2003, Statistical models for company growth, *Physica A: Statistical Mechanics and its Applications* 326, 241–255.